PRIMARY STRUCTURE OF PROTEIN S18 FROM THE SMALL *ESCHERICHIA COLI* RIBOSOMAL SUBUNIT

Makoto YAGUCHI*
Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Berlin-Dahlem, Germany

Received 16 September 1975

1. Introduction

Protein S18 is a very basic protein of the *E. coli* 30 S ribosomal subunit [1]. It has been implicated in aminoacyl-tRNA binding [2], and it appears to be situated at the decoding site of mRNA because the bromoacetylated AUG analog was found to be irreversibly bound to S18 [3,4]. It has a sulfhydryl group which reacts readily with *N*-ethylmaleimide and its uptake paralleled ethylmaleimide-induced loss of incorporating activity [5].

The tryptic peptides of protein S18 have been isolated and their amino acid compositions reported [6,7]. The N-terminal residue was found to be 'blocked' [7–9]. The structural gene for the protein S18 maps at minute 84 [7,10,11] which is outside of the major cluster for ribosomal protein genes around minute 64 [12]. An altered protein S18 was isolated from a temperature-sensitive mutant of *E. coli* and the alteration in the mutant protein consists of a replacement of arginine by cysteine [7].

This report summarizes the determination of the complete amino acid sequence of protein S18. More details will be published elsewhere. Protein S18 consists of 74 amino acids and the blocked N-terminal residue was found to be *N*-acetyl-alanine. The molecular weight of S18 is 8951.

* Visiting scientist. Permanent address: Division of Biological Sciences, National Research Council of Canada, Ottawa, Canada.

2. Materials and methods

Protein S18 was isolated from *E. coli* strain K as previously described [13] and provided by Dr H. G. Wittmann. The identity and purity of the protein was checked by two-dimensional polyacrylamide gel electrophoresis [14]. Tryptic and chymotryptic digestion was performed at pH 8.0 at 37°C for 4 or 20 h. Digestion with thermolysin was at 55°C and that with *Staphylococcus aureus* protease kindly supplied by Dr G. R. Drapeau, University of Montreal, was in 50 mM acetic acid, pH 4.0, for 16 h [15]. The isolation of the peptides (SP1, SP2, and SP3) from the *Staphylococcal* protease digest was achieved by gel filtration of the mixture (5–10 mg) on a Sephadex G-50 (fine) column (250 × 1.5 cm) at room temperature, and 15% acetic acid was used for the elution. Digestion with carboxypeptidase A and B was in 0.2 M *N*-ethylmorpholine buffer at pH 8.0 [16]. The 'finger-print' of various peptides was made on a sheet of thin-layer cellulose [17] and the peptides were extracted with 5.7 N HCl which contained 0.02% mercaptoethanol for analysis of their amino acid composition or with 30% acetic acid for preparative purposes. Amino acid analyses were performed with the single column procedure on a Durrum amino acid analyzer model D-500. Amino acid sequence of peptides was determined by the combined Dansyl-Edman technique [18,19]. Sequence determination of peptide SP2 was also performed in a solid-phase sequenator according to Laursen [20–22]. Automatic Edman-degradation of peptide SP3, after reacting with Braunitzer's reagent IV [23], was made in a liquid phase sequenator ac-

cording to Edman and Begg [24] with an improved Beckman sequenator [25,26].

3. Results and discussion

The protein S18 has an amino acid composition of Asp_3, Asn_1, Thr_6, Ser_3, Glu_2, Gln_5, Pro_2, Gly_3, Ala_7, Val_2, Ile_6, Leu_5, Tyr_6, Phe_3, His_1, Lys_6, Arg_{12}, Cys_1. Methionine and tryptophan are absent. The 19 basic amino acids and 5 acidic residues are comparible with the high isoelectric point of pH > 12 [27]. The mole percent of tyrosine (8.1%) is higher than that of other ribosomal proteins [28]. The number of amino acids obtained from the sequence data is an excellent agreement with the residues calculated from the amino acid composition of the protein.

Protein S18 was very effectively and specifically cleaved into 3 fragments (SP1, SP2, SP3) by the *Staphylococcal* proteinase at pH 4.0, and these fragments were isolated after gel filtration on a Sephadex G-50 column. The sum of the amino acid residues of the three peptides was exactly the same as the total amino acid residues of protein S18. The N-terminal residue of protein S18 and peptide SP1 was blocked. Glutamine and histidine were released by carboxypeptidase A from the C-terminal end of protein S18 and peptide SP3; thus the alignment of SP1–SP2–SP3, as shown in fig.1, was established.

Peptide SP1 was further degraded by trypsin, chymotrypsin and thermolysin. The N-terminal tryptic peptide T1, which is a dipeptide consisting of blocked alanine and arginine, was treated by carboxypeptidase B to remove arginine. The remaining blocked alanine residue was identified as *N*-acetyl-alanine by comparison with authentic *N*-acetyl-alanine (Bachem Feinchemikalien AG, Liestal, Switzerland) and *N*-formyl-alanine kindly provided by Dr B. Gutte, University of Köln, with the thin-layer chromatography. The blocked alanine and *N*-acetyl-alanine moved to the identical position (R_f=0.66). Other tryptic peptides were sequenced manually and chymotryptic and thermolytic peptides were used to align the tryptic peptides. Peptide SP2 was sequenced completely by the manual Dansyl-Edman technique and the automatic Edman degradation by the solid-phase sequenator. Both gave identical sequences. Peptide SP3 was directly sequenced by an improved Beckman sequenator. It was possible to identify the first 37 residues out of the total 40 residues by this sequenator run. Carboxypeptidase A produced glutamine and histidine, and addition of carboxypeptidase B to the same reaction mixture released arginine, aspartic acid and threonine. The manual Edman-degradation of the tryptic peptide T19 and the C-terminal chymotryptic peptide (70–74) established the sequence of the C-terminal region, thus completing the sequence of peptide SP3. Tryptic, chymotryptic and thermolytic peptides isolated directly from protein S18 can be fitted to the portions of the proposed sequence as shown in fig.1. Furthermore the tryptic peptides are manually sequenced to confirm the amino acid sequence.

Both SP1 and SP3 are very basic but SP2 has more acidic residues than basic one, and there are two basic clusters in the S18: position 5–8 (Arg–Arg–Arg–Lys) and position 59–62 (Lys–Arg–Ala–Arg). It is interesting to note that 7 out of total 9 aromatic residues are linked to lysine or arginine: Arg–Tyr (2–3; 62–62), Lys–Tyr (49–50), Tyr–Lys (23–24), Phe–Arg (4–5), Lys–Phe (8–9), Arg–Phe (11–12). The sequence Ala–Arg–Tyr is found in both positions 1–3 and 61–63.

Kahan et al. [7] reported that *E. coli* mutant 258ts has one more cysteine and one less arginine than the wild type and found a replacement of the arginine at the C-terminal position of a tryptic peptide Phe–Cys–Arg by cysteine in the mutant protein. From the sequence data presented in fig.1, it is concluded that the replacement site must be position 11.

One of the ribosomal protein (B-S19) from *Bacillus stearothermophilus* has been suggested to be homologous to *E. coli* E-S18 judging from its amino acid composition, molecular weight and two dimensional electrophoretic mobility [29,30], and the N-terminal amino acid sequence of B-S19 has been reported [9]. The comparison of the N-terminal amino acid sequence of *E. coli* S18 and *Bacillus stearothermophilus* S19 shows that they are homologous proteins and both cysteine and basic residues are conserved during evolution. These residues are probably important for function.

Both proteins have alanine as the N-terminal residue, but the alanine of E-S18 is acetylated while that of B-S19 is not acetylated suggesting the acetylation is not essential for the function. N-terminal acetyl-alanine was also found in cucumber virus 4, rabbit muscle enolase, wool, human erythrocyte carbonic anhydrase and frog hemoglobin [31]. Ribosomal protein L7 from *E.*

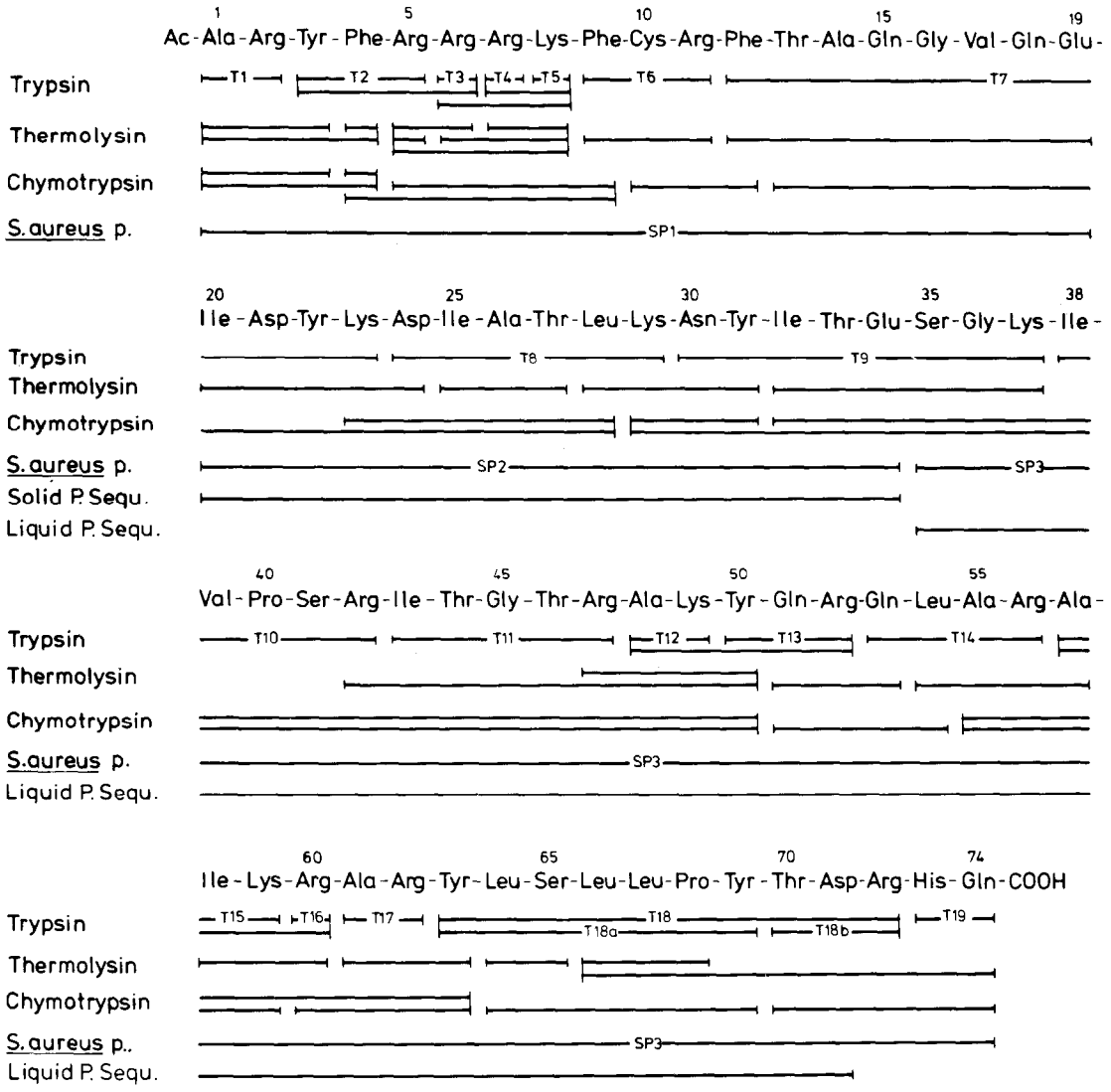

Fig.1. Amino acid sequence of protein S18 from *E. coli* ribosomes. Trypsin = tryptic peptides. S. aureus p. = peptides from digestion with *Staphylococcus aureus* protease. Solid p. sequ. = sequence elucidated by solid phase sequenator. Liquid p. sequ. = sequence elucidated by an improved Beckman sequenator.

coli has also a blocked N-terminal residue but it is acetyl-serine [32].

The secondary structure of S18 was predicted using the method of Chou and Fasman [33,34] and it is estimated to contain 31% α-helical region and 31% β-sheet structure. According to this method α-helical regions are positions 14–19 and 51–61, and β-sheet regions are 9–13, 20–33, and 43–46.

Acknowledgements

I thank Dr H. G. Wittman and Dr B. Wittmann-Liebold for many stimulating discussions, Dr E. Schiltz and Dr R. Chen for helpful suggestions, and Mrs J. Kuhlmeyer and Mrs J. Krauss for excellent technical assistance. I also thank Dr G. R. Drapeau for *Staphylococcus aureus* protease, and Dr B. Gutte for *N*-formyl-alanine.

References

[1] Wittmann, H. G. and Wittmann-Liebold, B. (1974) in: Ribosome, pp. 115–140 Cold Spring Harbor Monograph Series.
[2] Highland, J. H., Ochsner, E., Gordon, E., Hasenbank, R. and Stöffler, G. (1974) J. Mol. Biol. 86, 175–182.
[3] Pongs, O. and Lanka, E. (1975) Proc. Nat. Acad. Sci. USA 72, 1505–1509.
[4] Pongs, O. and Lanka, E. (1975) Hoppe Seyler's Z. Physiol. Chem. 356, 449–458.
[5] Moore, P. B. (1971) J. Mol. Biol. 60, 169–184.
[6] Wittmann-Liebold, B. (1971) Hoppe Seyler's Z. Physiol. Chem. 352, 1705–1714.
[7] Kahan, L., Zengel, J., Nomura, M., Bollen, A. and Herzog, A. (1973) J. Mol. Biol. 76, 473–483.
[8] Wittmann-Liebold, B. (1973) FEBS Lett. 36, 247–249.
[9] Yaguchi, M., Matheson, A. T. and Visentin, L. P. (1974) FEBS Lett. 46, 296–300.
[10] Bollen, A., Faelen, M., Lecocq, J. P., Herzog, A., Zengel, J., Kahan, L. and Nomura, M. (1973) J. Mol. Biol. 76, 463–472.
[11] DeWilde, M., Michel, F. and Bronman, K. (1974) Molec. gen. Genet. 133, 329–333.
[12] Jaskunas, S. R., Nomura, N. and Davies, J. (1974) in: Ribosomes, pp. 333–368 Cold Spring Harbor Monograph Series.
[13] Hindennach, I., Stöffler, G. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 7–11.
[14] Kaltschmidt, E. and Wittmann, H. G. (1970) Analyt. Biochem. 36, 401–412.
[15] Houmard, J. and Drapeau, G. R. (1972) Proc. Nat. Acad. Sci. USA 69, 3506–3509.
[16] Ambler, R. P. (1972) in: Methods in Enzymology, XXV, pp. 143–154 Academic Press.
[17] Bitar, K. G. (1974) Biochim. Biophys. Acta 386, 99–106.
[18] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[19] Bruton, C. J. and Hartley, B. S. (1970) J. Mol. Biol. 52, 165–178.
[20] Laursen, R. A. (1971) Eur. J. Biochem. 20, 89–102.
[21] Laursen, R. A., Horn, M. J. and Bonner, A. G. (1972) FEBS Lett. 21, 67–70.
[22] Wittmann-Liebold, B. and Lehmann, A. (1975) First International Conference on Solid-Phase Methods in Protein Sequence Analysis, Boston, Mass., in press.
[23] Braunitzer, G., Schrank, B., Ruhfus, A., Petersen, S. and Petersen, U. (1971) Hoppe-Seyler's Z. Physiol. Chem. 352, 1730–1732.
[24] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[25] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.
[26] Wittmann-Liebold, B., Geissler, A. W. and Marzinzig, E. (1975) manuscript submitted.
[27] Kaltschmidt, E. (1971) Analytical Biochem. 43, 25–31.
[28] Kaltschmidt, E., Dzionara, M. and Wittmann, H. G. (1970) Molec. gen. Genet. 199, 292–297.
[29] Isono, K., Isono, S., Stöffler, G., Visentin, L. P., Yaguchi, M. and Matheson, A. T. (1973) Molec. gen. Genet. 127, 191–195.
[30] Isono, S. and Isono, K. (1975) Eur. J. Biochem. 50, 483–488.
[31] Narita, K. (1970) in: Protein sequence determination, (S. B. Needleman, ed.) pp. 82–89, Springer-Verlag.
[32] Terhorst, C., Möller, W., Laursen, R. and Wittmann-Liebold, B. (1973) Eur. J. Biochem. 34, 138–152.
[33] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 211–222.
[34] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 222–245.